様式６

申請先学部　　　追加枠　採択番号　No.2

令和２年度大阪大学未来基金「学部学生による自主研究奨励事業」研究成果報告書

<table>
<tr><td>ふ　り　が　な
氏　　　名</td><td colspan="2">ご　そうせい
Wu Zengqing</td><td>学部
学科</td><td>人間科学部
人間科学コース</td><td>学年</td><td>２年</td></tr>
<tr><td rowspan="3">ふりがな
共　同
研究者氏名</td><td colspan="2"></td><td rowspan="3">学部
学科</td><td></td><td rowspan="3">学年</td><td>年</td></tr>
<tr><td colspan="2"></td><td></td><td>年</td></tr>
<tr><td colspan="2"></td><td></td><td>年</td></tr>
<tr><td colspan="2">アドバイザー教員
氏名</td><td>肖 川</td><td>所属</td><td colspan="3">情報科学研究科</td></tr>
<tr><td colspan="2">研究課題名</td><td colspan="5">The Impact of Grading Systems of Education on Equity, Teaching Efficiency, and Learning Strategies</td></tr>
<tr><td colspan="2">研究成果の概要</td><td colspan="5">研究目的、研究計画、研究方法、研究経過、研究成果等について記述すること。必要に応じて用紙を追加してもよい。（先行する研究を引用する場合は、「阪大生のためのアカデミックライティング入門」に従い、盗作剽窃にならないように引用部分を明示し文末に参考文献リストをつけること。）</td></tr>
</table>

Introduction

As the number of students going to international schools and studying abroad continues to grow globally, students' options in education system are becoming more diverse. Systems can have very different impacts on different students, making it increasingly important for each student and family to choose the most suitable learning system. Against this backdrop, high school academic systems such as A-level, IB, SAT, GAOKAO, and other teaching evaluation systems that are widely accepted by universities have significantly increased their impact on students. Although examinations such as SAT are not the only criteria used to evaluate admissions to the universities of the United States, these standard results still dominate the admissions process. It is necessary to study how education and grading systems affect the development of students with different socio-economic backgrounds and learning strategies and to assess what students can expect in terms of final assessment outcomes.

On the other hand, the education assessment system has a great impact not only on the development of students but also on institutions such as universities. Schools need to use ratings to select students for higher education or to allocate specific learning resources and form different learning communities within the institution. A well-balanced selection system needs to be equitable to enable the selectors to get better candidates, and the selectees are more likely to perform their true proficiency. However, it is difficult to have a system that can meet the conditions of all simultaneously (Weinberg, 1975), and there are such problems in evaluation systems of education. In the case of the IB and GAOKAO, for instance, the IB is a typical 7-point test, while GAOKAO is a 150-point test. Lower differentiation can lead to homogeneity in evaluation and difficulties in selection. A classic query about differentiation is why a 9-point difference between an 80 and an 89 can result in the same B grade, while a difference of only 1 point between a 79 and an 80 makes a qualitative difference (Tomar, 2019). The boundaries between grades are difficult to define precisely. In contrast to the problem of homogeneity, although the homogeneity of different scores seems to weaken the fairness of the grades, it should have a stronger motivational effect on students, especially those with average academic performance. To the extent that final performance is influenced by a normal distribution, a relatively large number of median grades will appear to present superior student performance overall. This produces a relatively stronger motivational effect, and ultimately educational outcomes may also be improved.

This paper examines the effects of grading systems on students and selection based on the concept of homogeneity and motivational capacity for learning. Homogeneity and motivational capacity, as the nature of the grading system of education, are a binary pair of complementary relationships. We examine their effects in different grading systems and whether there exists a preferable equilibrium. We also suggest options for learning environments for students with different categories of learning goals and motivations.

Preliminaries

How Motivation Affects Academic Performance. We mainly discuss the motivation of students based on self-determination theory, a widely used theory of motivation which examines the role of different types of motivation on the overall performance of motivation. The theory classifies motives into the six categories in order of their externalities from largest to smallest. The first three of these categories belong to extrinsic controlled motivation (CM); the last three belong to autonomous motivation (AM).

Intrinsic motivation and amotivation are highly positive predictors of eventual academic performance (Deci & Ryan, 1985; Breva & Galindo, 2020). Kusurkar et al. (2013) propose a correlation chain on how motivation affects final academic performance based on the relationship between AM and CM. Given that motivation, regardless of type, originates from the same entity, developing an indicator RAM to synthesize the roles of AM and CM to represent relative autonomous motivation could make the indicators more consistent. Research indicates that higher RAM is statistically associated with better learning strategies, which in turn further promote academic effort and ultimately act on academic performance. Mechanistically, learning strategies and learning effort are the mediating components in the relation between RAM and academic performance. Autonomous motivation and controlled motivation have a decisive influence on academic performance.

Method for Measuring Students' Motivation Levels. An important measure of student motivation levels is the use of the Academic Motivation Scale (AMS). Its distinction provides a referable level of differentiation for boundary-blurred motivational states.

Table 1. Calculated weights for subscale in the Academic Motivation Scale

Subscale	Motivation Type	Weight
IM knowledge, accomplishment, stimulation	AM	2
Identified regulation	AM	1
Introjected regulation	-	0
External regulation	CM	1
Amotivation	CM	2

Grade as a Primary Factor in Controlled Motivation. Regarding how grades specifically affect external motivation, Ackerman and Gross's study (2020) refers to a student's expectations of grades as an important predictor of that student's eventual academic performance, with the practical implication of indicating that we should not use a student's actual grades as a measure of external motivation, but rather the difference between a student's expectations of grades and actual grades. In addition, students are affected differently by their performance depending on their background conditions. Low-achieving students are more negatively affected by ratings and higher rubrics, while conversely, high-achieving students benefit, and this feedback is closely related to student expectations (Klapp, 2015).

Features of Grading System. The discrepancy between students' expectations and actual performance is not only related to their academic self-perceptions but also to the construction of the grading system itself. This is because different grading systems of education have different scales of scores and produce different levels of motivation. From the purpose of assessment, the function of evaluation is to control and enhance learning (Dahlgren et al., 2009). Whether it is to control the process of learning or to enhance motivation to learn, grading systems of education rely on performance as a mediating indicator.

Concerning the distribution of test scores, Deutsch (1979) notes in his paper that, in the vast majority of cases, scores of test-takers follow a normal distribution rather than reflecting their true levels of academic capability. This normal distribution can be achieved by setting the difficulty of test questions based on an estimate of student learning level and by discounting raw scores into the relevant grades. In a normal distribution, there is always a shortage of scores in the high scoring segments. The result is that in the more differentiated exams, the greater the gap between the performance of most candidates and the performance of the higher band.

Grading curves based on a normal distribution limit the allocation of overall high scores and are an artificial shortage; this is not a natural reflection of performance and therefore increases the potential for negative incentives for students (Deutsch, 1979). The prior academic performance serves as an incentive to influence students' performance the next time. Thus, the relative strengths in education of individuals are continually augmented. A student's background conditions and current learning accumulation will further influence the future distribution of performance as a consequence of assessment. In education, there is some fairness and efficiency in using a normal distribution for selection because a normal distribution allows for consistent distribution of performance across teachers and helps remove the inevitable effects of subjectivity and standard differences in grading. However, scores should relate only to learning standards. We should note that while it is true that scarcity exists in broad economic systems, motivation is not a scarce resource in the learning phase rather than the selection phase.

An Integrated Model of Learning Motivation. From the above discussion, we identify a pattern of how grading systems of education affect students' academic performance through grades: actual grades may differ from students' actual academic capability or expectations of performance, because of the normal distribution of the performance curve. This difference affects students' controlled motivation as an external motivator for learning, and is ultimately reflected in the outcome of the next examination.

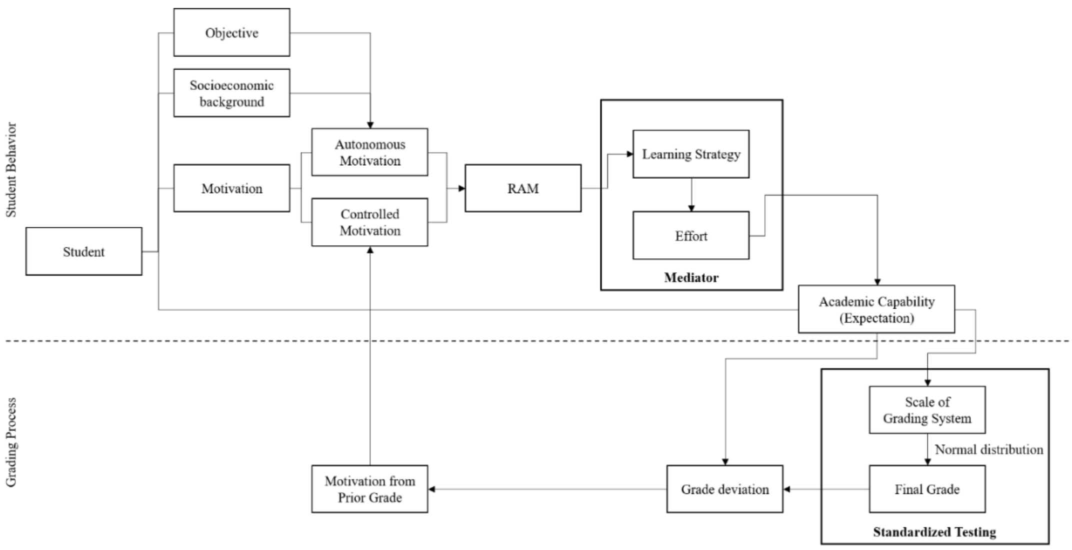

Fig 1. Iterative model of academic achievement motivating learning outcomes

Methodology

Iterative Model of Testing and Motivation Generation. We have developed an understanding of the interaction between testing and motivation, and based on this relationship we can construct a motivational system about student learning. A medium to the long term learning process will consist of several examinations, which we denote as a learning phase between two examinations. In this dynamic system, a learning phase consists of three steps.

1. Changing the current academic capability based on current learning motivation and learning progress. This can be represented by the experience formula:

$$\Delta Academic\ Capability = Effort \times Efficiency \tag{1}$$

2. A student's current academic capability, when compared to other students in the learning community on the exam, is converted into an actual grade under the scale of grading, with the score of all students in the community conforming to a normal distribution.
3. The difference between a student's actual score and performance based on the student's academic capability generates external motivation for the student.

The level of effort and efficiency in learning should have significant predictive ability for the increase in academic effort. At each stage of learning, the magnitude of effort is positively correlated with external motivation, and existing research suggests that the sigmoid function provides a better fit to motivation-effort data (Morris et al., 2020). The sigmoid function fitting to learning effort is controlled by two parameters, bias, and sensitivity, where Morris' study states that bias refers to the amount of motivation needed to initiate effort, while sensitivity to motivation determines the increment of motivation needed to raise effort. Thus, if a student's autonomous motivation is higher, the lower the initial motivation required for effort, the lower the bias and the more the function will shift to the left; autonomous motivation is less dependent on controlled motivation and therefore less sensitive to incentive, and the flatter the slope of the sigmoid function.

$$Effort(motivation) = t \times \frac{1}{1 + e^{\frac{-motivation - bias}{insensitivity}}} \tag{2}$$

As stated in our study, controlled motivation is mainly influenced by the difference between expected performance based on actual academic capability and actual performance after normal distribution. A student's expected performance can be obtained by standardizing the academic capability to the grading system's score interval. For an exam, its normal distribution of actual performance can be represented by $N(\mu, \sigma)$, where μ is the expected mean score for the exam and σ refers to standard deviation, and in the test sense it indicates the gradient of test difficulty. The greater the absolute value of the mean score set by the exam and the gradient of difficulty, the lower the actual score will be for students with relatively low actual academic capability. Differences in expected and actual performance within a learning community map onto CM and ultimately create different motivation for students.

Along with learning effort, another factor that affects the increase in learning ability is learning efficiency. The theory of the learning curve suggests that as the number of studies and attempts increases, the corresponding work becomes less costly and more efficient. The empirical curve is:

$$y = kx^{\frac{\log b}{\log 2}} \tag{3}$$

In the context of student learning, each of these variables can be specifically interpreted as: k is the time required to learn the basics as determined by the difficulty of the learning content, x is the total learning time spent on this learning content, and y is learning efficiency, representing the increase in academic capacity every y hours after x hours of effective learning. For ease of calculation, all-time units are given in hours. It is apparent from the nature of the function that the higher the learning rate b, the slower the improvement in learning efficiency. Several studies suggest that the learning rate for most learning should be in the interval $[0.75, 0.9]$ (Hax & Majluf, 1982).

Simulation Setup. The variables to be controlled in the simulation are the student conditions, the overall difficulty of the learning content, and the scales used in the test. The learning rate b, which indicates the difficulty of learning, takes the value $\{x|x = 0.01k, k \in [75,90]\ and\ k \in \mathrm{N}\}$ and the scale of grading takes the value $\{x|x \in [7,160]\ and\ x \in \mathrm{N}\}$. The student conditions are simulated based on different distributions, and all students' autonomous motivation and their basis of learning are randomly generated with three data distributions: positive skew, uniform, and negative skew distributions, respectively. All simulations are divided into $3 \times 3 = 9$ groups according to the distribution of student conditions. To ensure the generality of the generated data, for each group of simulations, we randomly generate 50 groups of student conditions and averaged them as the simulation results for the corresponding group. The number of overall simulations is $9 \times 50 \times 15 \times 154 = 1039500$ times.

Table 2. Parametric characteristics of the simulation

Group	Student Conditions Distribution		Learning Rate (b)	Scale
	AM	k		
1	Negative skew distribution	Negative skew	15 Categories	154 Categories
2		Uniform		
3		Positive skew		
4	Uniform distribution	Negative skew		
5		Uniform		
6		Positive skew		
7	Positive skew distribution	Negative skew		
8		Uniform		
9		Positive skew		

Summary of Results

After a massively repeated simulation, we first present the statistical results of the data on average academic capability in the form of trends. The trends in Table 3 are labeled with "↑", "↓", and "-" to indicate that the corresponding indicators increased, decreased, do not change significantly, or are irrelevant as the scale of the test increased.

Table 3. Trends in average academic capacity with the scale of tests

Group	Average Capability	High Capability Average	Low Capability Average
1	-	-	-
2	-	-	-
3	-	↑	-
4	-	↑	↓
5	↓	↑	↓
6	↓	↑	-

7	↑	↑	↓
8	-	↑	↓
9	↓	↑	↓

The overall trend is observed for each of the three indicators. In general, high levels of academic capability average increase with higher test scales and low capability average decrease with higher scales, while average capability is related to the specific conditions of the students. This suggests that the greater the test differentiation, the greater the order of magnitude of the motivation for students, and the greater the variance in the context of a competitive system.

In terms of magnitude, as the test scale rises, the difference in the improvement of the high capability average is smaller, while the low capability average decreases to a significant degree. The statistical results also show that an increase in the value of the learning rate only leads to an absolute decrease in the learning capability, but has no effect on the pattern of the relationship between average learning capability and the scale of examinations. In particular, the higher the learning rate, the greater the learning difficulty, and the more pronounced the increase in the average score of the high band with the examination scale.

Conclusion

1. Test scales have a greater effect on the balance between education efficiency and education equity. The greater the differentiation in the examination scale, the more conducive it is to produce high capability students and the more efficient the teaching, but it has a greater negative impact on students with a lower learning base or autonomous motivation, and the less equitable the education. The basis of learning and autonomous motivation is related to the objective of students as well as their socioeconomic background.
   From the perspective of a steady progression of students, students with low initial goals or who are less able to adapt to learning should avoid studying in the more differentiated assessment system. Students who have high goals and a good learning foundation can participate in the more differentiated examinations. For students with strong self-motivation but a weak learning base, a lower-scaled test is inappropriate. Concerning goal setting, since motivation is based on expectations, it is more helpful to set short-term goals that are consistent with the individual's ability to learn.
   In addition, from the perspective of balancing efficiency and fairness in education, the schools could choose to combine lower and higher differentiated tests, using lower differentiated score systems in the long-term learning process and higher differentiated score systems in the selection process, which can enhance the equity of the output of teaching and learning while ensuring the rationality of the selection process.
2. Concerning the selection of appropriate learning community based on learning interests, in general, students perform better in groups where there is a general willingness to learn autonomously, independent of differentiation and external incentives; the incentives in randomly formed learning communities are contingent, and overall, such learning environments are more conducive to the progress of the best students; and motivation in groups with low autonomous motivation are susceptible to controlled motivation. Accordingly, it is important to choose an assessment system with a relatively low level of differentiation for the students with low self-motivation. From this, we can also understand that maximizing students' interest in learning is of great practical importance for learning.

Table 4 is a comprehensive recommendation based on the above data as to what type of test scale and learning environment is appropriate for varied types of students.

Table 4. Matching of test scales and learning environments to student types

Type of Student		Scale	AM of Learning Community
Basis of Learning	Objective / AM		
High	High	High	High or Uniform
High	Low	High or Medium	High or Uniform or Low
Low	High	Low	High or Low
Low	Low	Low	Low

The goals of system optimization and the environment are not always aligned, therefore, students and parents can select an appropriate grading system of education based on benchmark values which may result in a more coherent and effective status of the education system generally.

Reference

Ackerman, D. S., & Gross, B. L. (2020). You gave me a B-?! Self-Efficacy, implicit theories, and student reactions to grades. Journal of Marketing Education, 42(2), 149-156.

Breva, A., & Galindo, M. P. (2020). Types of motivation and eudemonic well-being as predictors of academic outcomes in first-year students: A self-determination theory approach. PsyCh Journal.

Dahlgren, L. O., Fejes, A., Abrandt-Dahlgren, M., & Trowald, N. (2009). Grading systems, features of assessment and students' approaches to learning. Teaching in Higher Education, 14(2), 185-194.

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. New York, NY: Plenum.

Deutsch, M. (1979). Education and distributive justice: Some reflections on grading systems. American Psychologist, 34(5), 391.

Hax, A. C., & Majluf, N. S. (1982). Competitive cost dynamics: the experience curve. Interfaces, 12(5), 50-61.

Klapp, A. (2015). Does grading affect educational attainment? A longitudinal study. Assessment in Education: Principles, Policy & Practice, 22(3), 302-323.

Kusurkar, R. A., Ten Cate, T. J., Vos, C. M. P., Westers, P., & Croiset, G. (2013). How motivation affects academic performance: a structural equation modelling analysis. Advances in health sciences education, 18(1), 57-69

Morris, L. S., Norbury, A., Smith, D. A., Harrison, N. A., Voon, V., & Murrough, J. W. (2020). Dissociating self-generated volition from externally-generated motivation. PloS one, 15(5), e0232949.

Tomar, D. A. (2019). Eliminating the grading system in college: The pros and cons. The Quad Retieved, Jan.

Weinberg, G. M. (1975). An introduction to general systems thinking (Vol. 304). New York: Wiley.

Remark

This research is submitted to the journal *Computers & Education* (Elsevier) and is under review.